



**PCT**  
WELTORGANISATION FÜR GEISTIGES EIGENTUM  
Internationales Büro  
INTERNATIONALE ANMELDUNG VERÖFFENTLICHT NACH DEM VERTRAG ÜBER DIE  
INTERNATIONALE ZUSAMMENARBEIT AUF DEM GEBIET DES PATENTWESENS (PCT)

<b>(51) Internationale Patentklassifikation <sup>6</sup> :</b> <b>G06F 1/00, G07C 9/00</b>	<b>A2</b>	<b>(11) Internationale Veröffentlichungsnummer: WO 98/06020</b> <b>(43) Internationales Veröffentlichungsdatum:</b> 12. Februar 1998 (12.02.98)
<b>(21) Internationales Aktenzeichen:</b> PCT/EP97/04221 <b>(22) Internationales Anmeldedatum:</b> 2. August 1997 (02.08.97)  <b>(30) Prioritätsdaten:</b> 196 31 484.4      3. August 1996 (03.08.96)      DE  <b>(71)(72) Anmelder und Erfinder:</b> BARTMANN, Dieter [DE/DE]; Taufkirchner Strasse 16, D-85435 Erding (DE).  <b>(74) Anwalt:</b> SEGETH, Wolfgang; Louis, Pöhlau, Lohrentz & Segeth, Postfach 30 55, D-90014 Nürnberg (DE).		<b>(81) Bestimmungsstaaten:</b> AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DK, EE, ES, FI, GB, GE, GH, HU, IL, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, US, UZ, VN, YU, ZW, ARIPO Patent (GH, KE, LS, MW, SD, SZ, UG, ZW), eurasisches Patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), europäisches Patent (AT, BE, CH, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI Patent (BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, NE, SN, TD, TG).  <b>Veröffentlicht</b> <i>Ohne internationalen Recherchenbericht und erneut zu veröffentlichen nach Erhalt des Berichts.</i>
<b>(54) Title:</b> METHOD FOR VERIFYING THE IDENTITY OF A USER OF A DATA PROCESSING UNIT WITH A KEYBOARD DESIGNED TO PRODUCE ALPHANUMERIC CHARACTERS  <b>(54) Bezeichnung:</b> VERFAHREN ZUR VERIFIZIERUNG DER IDENTITÄT EINES BENUTZERS EINER MIT EINER TASTATUR ZUR ERZEUGUNG ALPHANUMERISCHER ZEICHEN ZU BEDIENENDEN DATENVERARBEITUNGSANLAGE  <b>(57) Abstract</b>  A proposed method for verifying the identity of a user of a data processing unit with a keyboard designed to produce alphanumeric characters and involving at least one of the identification parameters derived from a model of actual observation of the typing behaviour of a user compared with a reference model parameter contained in a computer and determined from a sequence of events such as the striking or release of a key.  <b>(57) Zusammenfassung</b>  Vorgeschlagen wird ein Verfahren zur Verifizierung der Identität eines Benutzers einer mit einer Tastatur zur Erzeugung alphanumerischer Zeichen zu bedienenden Datenverarbeitungsanlage, bei dem aus dem Muster von aktuellen Beobachtungen an der Tastatur mindestens eine auf das Tippverhalten des Benutzers hinweisende Kenngröße abgeleitet und diese Kenngröße mit der Kenngröße verglichen wird, die aus einem in der Datenverarbeitungsanlage abgelegten Referenzmuster abgeleitet ist, wobei die Kenngröße aus der Abfolge der Ereignisse, Drücken oder Loslassen einer Taste, ermittelt wird.		

# **LEDIGLICH ZUR INFORMATION**

Codes zur Identifizierung von PCT-Vertragsstaaten auf den Kopfbögen der Schriften, die internationale Anmeldungen gemäss dem PCT veröffentlichen.

AL	Albanien	ES	Spanien	LS	Lesotho	SI	Slowenien
AM	Armenien	FI	Finnland	LT	Litauen	SK	Slowakei
AT	Österreich	FR	Frankreich	LU	Luxemburg	SN	Senegal
AU	Australien	GA	Gabun	LV	Letland	SZ	Swasiland
AZ	Aserbaidschan	GB	Vereinigtes Königreich	MC	Monaco	TD	Tschad
BA	Bosnien-Herzegowina	GE	Georgien	MD	Republik Moldau	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagaskar	TJ	Tadschikistan
BE	Belgien	GN	Guinea	MK	Die ehemalige jugoslawische Republik Mazedonien	TM	Turkmenistan
BF	Burkina Faso	GR	Griechenland	ML	Mali	TR	Türkei
BG	Bulgarien	HU	Ungarn	MN	Mongolei	TT	Trinidad und Tobago
BJ	Benin	IE	Irland	MR	Mauretanien	UA	Ukraine
BR	Brasilien	IL	Israel	MW	Malawi	UG	Uganda
BY	Belarus	IS	Island	MX	Mexiko	US	Vereinigte Staaten von Amerika
CA	Kanada	IT	Italien	NE	Niger	UZ	Usbekistan
CF	Zentralafrikanische Republik	JP	Japan	NL	Niederlande	VN	Vietnam
CG	Kongo	KE	Kenia	NO	Norwegen	YU	Jugoslawien
CH	Schweiz	KG	Kirgisistan	NZ	Neuseeland	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Demokratische Volksrepublik Korea	PL	Polen		
CM	Kamerun	KR	Republik Korea	PT	Portugal		
CN	China	KZ	Kasachstan	RO	Rumänien		
CU	Kuba	LC	St. Lucia	RU	Russische Föderation		
CZ	Tschechische Republik	LI	Liechtenstein	SD	Sudan		
DE	Deutschland	LK	Sri Lanka	SE	Schweden		
DK	Dänemark	LR	Liberia	SG	Singapur		
EE	Estland						

5     Verfahren zur Verifizierung der Identität eines Benutzers  
      einer mit einer Tastatur zur Erzeugung alphanumerischer  
      Zeichen zu bedienenden Datenverarbeitungsanlage

10     Problembeschreibung

Ein unbefugter Zugriff auf Daten, die in einem Rechner  
abgelegt sind, die unerlaubte Benutzung von  
Computerprogrammen, eine unsichere oder fehlende  
Identitätsprüfung beim Zugang zu Online Services oder bei  
15   Interactive TV, die mißbräuchliche Verwendung von  
magnetstreifen- bzw. chipbehafteten Bankkarten sowie das  
Fehlen einer nachweisbaren Zuordnung eines am Computer  
erstellten Datenobjektes zum Ersteller, können zu großen  
wirtschaftlichen und gesellschaftlichen Schäden führen.  
20   Dasselbe gilt für den Zutritt zu geschützten Bereichen in  
Gebäuden. Die Benutzeridentifikation am Computer ist  
deshalb sehr wichtig.

Sie geschieht in den weitaus meisten Fällen dadurch, daß  
25   der Benutzer ein vereinbartes Codewort (Paßwort,  
persönliche Identifikationsnummer, Benutzernummer etc.)  
angeben muß. Dieses Codewort wird von mir im folgenden  
stets Paßwort genannt. Stimmt dieses Wort, wird der  
Benutzer vom System als derjenige identifiziert, für den er  
30   sich ausgibt. Der damit erzielbare Schutz steht und fällt  
mit der Geheimhaltung bzw. Nicht-Erratbarkeit des  
Paßwortes.

Damit ein Paßwort nicht leicht erraten werden kann, muß es  
35   einigermaßen kompliziert sein und häufig gewechselt werden.

Dies stößt jedoch auf Widerwillen beim Benutzer, denn er sollte es auswendig wissen, oder, falls er es doch aufgezeichnet hat, diese Aufzeichnung einerseits schnell zugänglich und andererseits für jede andere Person  
5 unzugänglich aufbewahren.

Hier mangelt es am Arbeitsplatz oft an der nötigen Sorgfalt. Die praktische Erfahrung zeigt, daß sich eine hundertprozentige Geheimhaltung des Paßwortes nicht durchsetzen läßt. Dies gilt ebenso im privaten Bereich.  
10 Eine 1995 durchgeführte Umfrage des Emnid-Instituts hat ergeben, daß in über zehn Prozent aller deutschen Haushalte die zur Magnetstreifen- bzw. Chipkarte gehörige PIN (persönliche Identifizierungsnummer) mehr als einem Familienmitglied bekannt ist.

15 Das Paßwort wirkt darüber hinaus höchstens während des Zeitraum seiner Eingabe benutzeridentifizierend. Anschließend kann der Benutzer wechseln, ohne daß dies vom Computersystem bemerkt wird. Es ist auch möglich, den  
20 Benutzer zu einer unfreiwilligen Freigabe des Paßwortes zu bewegen.

Insgesamt genügt der Paßwortschutz wegen wegen seiner mangelnden Praktikabilität, wegen der ungenügenden  
25 Geheimhaltung und wegen der nur auf einen Zeitpunkt bezogenen Überprüfung den Sicherheitsbedürfnissen nur unzureichend.

30 Modernere Benutzeridentifikationssysteme arbeiten mit signifikanten biometrischen Merkmalen, z.B. dem Fingerabdruck, dem Augenhintergrund oder der Stimme. Sie besitzen aber ebenfalls spezifische Nachteile, so daß sie nur in Nischen zu Anwendung kommen.

Aufgabe

Der Erfindung liegt die Aufgabe zugrunde, ein  
Personenidentifizierungssystem zu entwickeln, welches  
5 gegenüber den bekannten Systemen folgende Vorteile besitzt:  
Es soll

- \* der Person selbst unbekannt und deshalb nicht so anfällig  
gegenüber Geheimhaltungsverlust sein wie der Paßwort-  
schutz
- 10 \* nur mit extrem großem Aufwand nachahmbar sein
- \* eine permanente Benutzeridentifizierung ermöglichen
- \* einen hohen Grad an Identifikationssicherheit gewähr-  
leisten (hohe personelle Signifikanz)

15 Stand der Technik

In diesem Kapitel wird lediglich informell erläutert, wie  
der derzeitige Stand der Technik ist und welche neuen Ideen  
dem hier vorgestellten System zu Grunde liegen. Eine  
20 genauere Betrachtung der Unterschiede findet sich in dem  
Kapitel "Exaktere Beschreibung der Unterschiede zum Stand  
der Technik" weiter unten.

Methode der Schreibdynamik

25

Aus der Vergangenheit sind bereits mehrere Systeme bekannt,  
die ebenfalls einen Tastaturbenutzer anhand seines  
Tippverhaltens identifizieren. Die den hier vorgestellten  
Ansätzen am nächsten liegende Methode stammt von S. J.  
30 Shepherd aus dem Jahr 1995 (in "European Convention on  
Security and Detection; Conf. Publ. No. 408; London: Inst.  
of Electrical Engineers; 1995; Seiten 111-114). Diese  
basiert auf der Schreibdynamik und dem Tastendruck. Erstere  
teilt sich dabei auf in die einzelnen Aspekte Übergangs-

und Anschlagdauer der einzelnen Tasten, sowie Tipprate und Fehlerfrequenz.

### Nachteile

5

An der Technischen Universität München wurde ein Programm entwickelt, das nach der oben genannten Methode arbeitet (Bartmann, Dieter: "Identifikation eines Tastaturbenutzers durch Analyse des Tippverhaltens". Diplomarbeit, Institut für Informatik der TU München, 1995). Ausführliche Tests haben gezeigt, daß diese Methode nicht genügend trennscharf ist. Die beiden Fehlerraten "False Rejection" und "False Acceptance" sind für eine praktische Anwendung zu groß. Dabei wurden im Programm wesentlich bessere statistische Verfahren herangezogen, als in der oben angeführten Schrift vorgeschlagen sind.

10  
15

Hauptgründe für das schlechte Abschneiden:

1. Die Schreibdynamik einer Person schwankt stark. Sie hängt von der Tagesform und weiteren exogenen Einflüssen ab. Das Programm muß diese Schwankungen tolerieren, um die False Rejection Rate unterhalb eines akzeptablen Wertes zu halten. Dadurch verliert die Methode wesentlich an Trennschärfe.
2. Handelsübliche Tastaturen besitzen ein sehr grobes zeitliches Auflösungsvermögen. Deshalb sind die gemessenen Übergangszeiten so stark gerundet, daß wesentliche Information verloren geht. Das zeitliche Auflösungsvermögen liegt bei schnell schreibenden Personen in der Größenordnung der Standardabweichung der Verteilung der Übergangszeiten.

20

25

30

Aufgabe angesichts dieses Standes der Technik

Neben der Schreibdynamik sind noch zusätzliche  
personentypische Merkmale zu finden, die über die Zeit  
5 hinweg wesentlich stabiler bleiben.

Idee

Es sind zur Benutzeridentifizierung zeitunabhängige  
10 Merkmale heranzuziehen. Dies hat den Vorteil, daß das  
Problem der mangelnden Zeitauflösung der Tastatur nicht  
mehr auftritt.

Als zeitunabhängige Merkmale werden herangezogen:

- 15 \* Die Tendenz des Benutzers, sich zu verhaspeln. Dies  
drückt sich unter anderem dadurch aus, daß der zweite  
Anschlag erfolgt, noch bevor der erste beendet ist  
(sog. Verschränkung oder Überholungen, siehe unten).  
Dabei handelt es sich nicht um Schreibfehler und ist  
auf dem Bildschirm auch nicht erkennbar. Es hat sich  
20 gezeigt, daß dieses Verhalten sehr grundlegend ist.
- \* Das Auswahlverhalten, falls mehrere Tasten zur  
Verfügung stehen (in erstere Linie Gebrauch der  
Shifttasten). Dieses Verhalten ist entscheidend dadurch  
geprägt, auf welche Weise eine Person das Schreiben auf  
25 der Tastatur gelernt hat und in welcher Handstellung  
sie schreibt (z.B. Auflage der Handballen auf dem  
Tisch). Auch dieses Merkmal ist äußerst konstant.
- \* Daneben gibt es noch weitere zeitunabhängige Merkmale,  
wie z.B. die Häufigkeit von Buchstabendrehern u.s.w.  
30 (siehe unten).

Ein Test mit echten Daten hat gezeigt, daß unter  
geschickter Heranziehung dieser personentypischen Merkmale  
die False Acceptance Rate weit unter ein Promille gedrückt

werden kann. Dabei wurden über 100.000 Angriffsversuche durchgeführt.

5 Aus den angeführten Gründen geht der hier erhobene Anspruch weit über den Stand der Technik hinaus.

### LÖSUNG

#### 10 Anwendungsgebiet

Die grundlegende Idee besteht darin, daß zum einen jeder Mensch eine ihm eigene, ihn charakterisierende Art zu tippen hat, und zum anderen dieses Tippverhalten durch eine  
15 technische Vorrichtung meßbar und analysierbar ist. Neben der Analyse des Tippverhaltens kann außerdem noch die Analyse des Umgangs mit der Maus oder anderer Zeigergeräte zur Identifizierung einer Person herangezogen werden. Die gesamte Methodik der Messung und Analyse kann dabei  
20 größtenteils übertragen werden. Da das Tippverhalten sicherlich am meisten personenidentifizierende Informationen enthält, ist die weitere Beschreibung hauptsächlich darauf ausgerichtet.

25 Für ein Verfahren, das die obige Idee in geeigneter Weise umsetzt, sind etliche Anwendungsmöglichkeiten denkbar:

- \* Identitätsprüfung, z.B. bei der Zutritts-  
/Zugriffskontrolle
- 30 Will eine Person in geschützte Bereiche eintreten oder auf geschützte Systeme zugreifen, so muß sie sich in der Regel bei dem Überwachungssystem unter einer ihr (der Person) zugeordneten Kennung anmelden. Anschließend muß sie sich anhand eines bestimmten



- Merkmals identifizieren, d.h. sie muß das System davon überzeugen, daß sie tatsächlich die zu der angegebenen Kennung gehörige Person ist. Derzeit übliche Verfahren hierzu sind die Eingabe von Paßwörtern bzw. PINs, die
- 5 Analyse des Fingerabdrucks, des Augenhintergrunds, der Stimme, des Gesichts sowie die Überprüfung von mit der Person verbundenen Gegenständen (Chipkarte/Magnetkarte/Schlüssel/...). Die hier vorgestellte Identifizierung durch Analyse des
- 10 Tippverhaltens einer Person könnte die bisherigen Verfahren ergänzen bzw. ersetzen.
- \* Authentifizierung eines elektronisch erstellten Dokuments
- 15 Erstellt jemand unter zuhilfenahme einer Tastatur ein Dokument, so kann während des gesamten Erstellungsvorgangs das Tippverhalten dieser Person mit aufgezeichnet werden. Werden diese gesamten Meßdaten untrennbar und unverfälschbar zusammen mit dem Dokument
- 20 abgespeichert, so ist dadurch die Identität des Schreibers direkt mit diesem verbunden. Es kann somit jederzeit überprüft werden, von wem ein derartiges Dokument tatsächlich stammt.
- 25 \* laufende Überprüfung der Identität eines Tastaturbenutzers
- Bei vielen Systemen fallen während der Arbeit mit ihnen immer wieder Tastatureingaben an. So müssen beispielsweise bei herkömmlichen Computern die Befehle,
- 30 der Text etc. zumindest teilweise über die Tastatur eingegeben werden. In solch einem Fall kann das Tippverhalten laufend analysiert und somit die Identität des Benutzers laufend überprüft werden. Dadurch könnten unberechtigte Zugriffe einer anderen

Person während der Abwesenheit der aktuell im System angemeldeten Person erkannt und verhindert werden.

- 5       \*   Identifizierung einer unauthorisiert handelnden Person  
Versucht eine Person sich unberechtigt unter der  
Kennung einer anderen Person bei einem System  
anzumelden, so kann das System (falls es den Angriff  
erkennt) versuchen, durch Analyse des Tippverhaltens  
des Angreifers dessen wahre Identität herauszufinden.

10

\* ...

- Der hier und auch im folgenden immer wieder erwähnte  
Begriff "Tastatur" bezieht sich nicht nur auf gewöhnliche  
15   Computertastaturen. Vielmehr sei damit ganz allgemein ein  
Gerät bezeichnet, mit dessen Hilfe eine Person durch  
Drücken einer oder mehrerer Tasten oder durch Berühren  
eines oder mehrerer Tastenfelder Eingaben tätigt.

20   Einführendes Beispiel

- Im weiteren soll die Methodik der Messung und der Analyse  
des Tippverhaltens näher erläutert werden. Dies soll anhand  
des folgenden Anwendungsbeispiels erfolgen. Die gesamte  
25   Vorgehensweise kann dann leicht auf alle möglichen anderen  
Anwendungen und zugrundeliegenden Systeme übertragen  
werden.

Beispiel

30

Ein Computersystem wird von mehreren Personen benutzt, die  
Eingaben in das System erfolgen zumindest u.a. über eine  
Tastatur. Bevor man mit dem System arbeiten kann, muß man  
sich bei ihm unter einer Kennung anmelden. Jeder Person ist

genau eine Kennung zugeordnet, mit der auch die jeweiligen Rechte des Benutzers verbunden sind. Nach der Eingabe der Kennung muß sich der Benutzer dem System gegenüber identifizieren. Dies soll mit Hilfe des personentypischen Merkmals "Tippverhalten" erfolgen.

### Lernvorgang

Damit das System die einzelnen Benutzer auf diese Art überhaupt identifizieren kann, muß ihm zunächst einmal bekannt gemacht werden, wie die verschiedenen Personen tippen. Dies geschieht in einer sogenannten Lernphase, die in unserem Beispiel wie folgt ablaufen soll:

Eine vom System bereits identifizierte Person, die die entsprechenden Berechtigungen besitzt, weist das System an, das Tippverhalten eines bestimmten Benutzers X zu lernen. Dazu muß dieser Benutzer X einen vom System auf dem Bildschirm ausgegebenen längeren Text (in der Regel mehrere Absätze lang) abtippen. Das System führt dabei geeignete Messungen durch und wertet schließlich die gemachten Beobachtungen in der weiter unten beschriebenen Weise aus. So erhält es ein für X typisches Referenzmuster. Mit dessen Hilfe kann in zukünftigen Indetifizierungsvorgängen dieser Benutzer immer wieder an seinem Tippverhalten vom System erkannt werden. Das Referenzmuster wird vom System in Verbindung mit der Kennung der Person X dauerhaft abgespeichert. Nach Abschluß der Lervorgänge für alle Benutzer liegt somit dem System zu jeder Kennung genau ein Referenzmuster, und zwar das des zugehörigen Benutzers, vor.

### Identifizierung

Bei jeder Anmeldung am System muß sich ein Benutzer Y - wie

oben bereits erwähnt - dem System gegenüber identifizieren. Dies soll in dem hier dargelegten Beispiel wie folgt vor sich gehen:

- 5      Nachdem der Benutzer Y die Kennung eingegeben hat, unter der er sich beim System anmelden will, wird er dazu aufgefordert, einen kurzen, auf dem Bildschirm angezeigten Text (in der Regel wenige Zeilen lang) abzutippen. Dabei führt das System wieder entsprechende Messungen durch. Die  
10     so gewonnenen Beobachtungen vergleicht es dann mit dem zu der angegebenen Kennung gehörigen Referenzmuster M und trifft auf der Basis dieses Vergleichs eine der beiden möglichen Entscheidungen: Liegt eine deutliche  
15     Übereinstimmung zwischen dem gerade beobachteten Tippverhalten von Y und dem im Referenzmuster M niedergelegten Verhalten vor, so wird Y vom System als die zur angegebenen Kennung gehörige Person identifiziert. Andernfalls bricht das System den Anmeldevorgang ab.

20     Verallgemeinerungen

- Das hier angeführte Beispiel stellt nur eine sehr einfache von vielen denkbaren Möglichkeiten dar, die zugrundeliegende Idee "Identifizierung einer Person anhand  
25     ihres Tippverhaltens" in einem System umzusetzen. Insbesondere können der Lernvorgang und die Identifizierung in vielerlei Hinsicht abgewandelt werden. So sind z.B. die folgenden Veränderungen denkbar:

- 30     \*    kein oder nur teilweise vorgegebener Text; der Rest muß als Freitext vom Benutzer eingetippt werden  
     \*    Aufteilung des Lernvorgangs auf mehrere "Sitzungen"  
     \*    laufende Messungen bei allen Tastatureingaben und damit

verbunden laufende Überprüfung der Identität eines Benutzers

- \* Einführung unterschiedlicher Sicherheitsstufen (gleich nach der Anmeldung dürfen nur unkritische Aktionen vorgenommen werden, der Zugriff auf sensitivere Befehle und Daten wird erst dann gestattet, wenn der Benutzer anhand der laufenden Eingaben auf einem sehr hohen Sicherheitsniveau identifiziert werden konnte)

\* ...

## Grundlagen

### Definitionen

- 15 In diesem Abschnitt sollen einige für das Weitere wichtige Begriffe definiert werden.

### Lerntext

- 20 Dies ist der von einem Benutzer während des Lernvorgangs eingegebene Text. In dem hier verwendeten Beispiel ist dieser Text gesamt vom System vorgegeben.

### Testtext

- 25 Dies ist der von einem Benutzer während eines Identifizierungsvorgangs eingegebene Text. In dem hier verwendeten Beispiel ist dieser Text gesamt vom System vorgegeben. Er kann für verschiedene
- 30 Identifizierungsvorgänge unterschiedlich sein.

Anmelder

5 Mit Anmelder wird diejenige Person bezeichnet, die sich bei dem System anmeldet. Insbesondere werden die im Verlauf der zugehörigen Identifizierung auftretenden Tastenanschläge alle vom Anmelder vorgenommen.

Behaupteter

10 Jedem Benutzer des Systems ist genau eine Kennung zugeordnet. Meldet sich eine Person beim System an, so muß sie dazu eine Kennung eingeben. Sie behauptet dadurch dem System gegenüber, die mit dieser Kennung verbundene Person zu sein.

15 Mit Behaupteter wird im folgenden immer diejenige Person bezeichnet, die zu der bei einem Anmeldevorgang eingegebenen Kennung gehört. Diese muß nicht zwingen mit derjenigen Person identisch sein, die die Kennung eingegeben hat.

20

Berechtigter

Stimmen bei einem Anmeldevorgang Anmelder und Behaupteter überein, so wird der entsprechende Benutzer auch als

25 Berechtigter bezeichnet. Er hat sich also unter seiner eigenen Kennung angemeldet.

Unberechtigter

30 Stimmen bei einem Anmeldevorgang Anmelder und Behaupteter nicht überein, so wird der Anmelder auch als Unberechtigter bezeichnet. Er hat sich also unter einer fremden Kennung angemeldet.

Gemessene Daten

Dieser Abschnitt erläutert die während des Lernvorgangs und während der Identifizierung vorgenommenen Messungen anhand der beispielhaften Eingabe des Wortes "Hund".

Die Eingabe von "Hund" kann durch die folgenden Aktionen geschehen (es sind auch andere Möglichkeiten denkbar):

1. Drücken der Taste "linkes\_Shift"
2. Drücken der Taste "h"
- 10 3. Loslassen der Taste "h"
4. Loslassen der Taste "linkes\_Shift"
5. Drücken der Taste "u"
6. Loslassen der Taste "u"
7. Drücken der Taste "n"
- 15 8. Loslassen der Taste "n"
9. Drücken der Taste "d"
10. Loslassen der Taste "d"

Um die weiteren Erläuterungen übersichtlicher gestalten zu können, wird die folgende symbolische Schreibweise eingeführt: Das Drücken einer Taste wird durch Taste↓, das Loslassen einer Taste durch Taste↑ dargestellt. Die oben angeführten zehn Ereignisse lassen sich damit kurz wie folgt schreiben:

linkes\_Shift↓ h↓ h↑ linkes\_Shift↑ u↓ u↑ n↓ n↑  
d↓ d↑

Obige Eingabe des Wortes "Hund" ist durch die zeitliche Abfolge dieser zehn Elementarereignisse charakterisiert. Als Elementarereignis soll im folgenden immer das Drücken oder das Loslassen einer beliebigen Taste verstanden werden. Für die Analyse des Tippverhaltens einer Person genügt es, die zeitliche Abfolge der von ihr hervorgerufenen Elementarereignisse zu erfassen. Das bedeutet, daß zu jedem Elementarereignis der Zeitpunkt

dessen Auftretens gemessen werden muß.

Die gesamte Information läßt sich somit durch die folgenden drei Meßgrößen erfassen:

- 5     1. Ereignisart:    { ↓, ↑ } als Wertebereich  
       Die Ereignisart zeigt an, ob eine Taste gedrückt oder  
       losgelassen wurde
2. Taste:       { a, b, c, ..., linkes\_Shift, rechtes\_Shift,  
       ... } als Wertebereich (z.B. 102 verschiedene Tastenbe-  
       zeichnungen)  
       Die Meßgröße Taste zeigt an, welche Taste gedrückt bzw.  
       losgelassen wurde
- 10     3. Zeitpunkt:     positive reelle Zahlen als Wertebereich  
       Der Zeitpunkt gibt an, zu welchem Zeitpunkt die Taste  
       gedrückt bzw. losgelassen wurde
- 15

Jedes Elementarereignis führt zu einer eigenen Beobachtung  
 (einer Messung dieser drei Meßgrößen). Die gesamte Menge  
 aller Beobachtungen, die während des Tippens eines gewissen  
 20 Textes (z.B. des Lern- oder des Testtextes) gemacht werden,  
 bilden das sogenannte Beobachtungsfeld dieses Tippvorgangs.  
 In unserem Beispiel könnte dies etwa wie folgt aussehen:

25

Ereignisart	↓	↓	↑	↑	↓	↑	↓	↑	↓	↑
Taste	linkes_Shift	h	h	linkes_Shift	u	u	n	n	d	d
Zeitpunkt	0,000	0,024	0,257	0,301	0,388	0,522	0,621	0,698	0,984	1,131

- 30     Alle weiteren Berechnungen und Analysen werden auf der  
       Basis derartiger Beobachtungsfelder durchgeführt.



Erweiterung

Zusätzlich zu den hier angegebenen Meßgrößen kann gegebenenfalls noch der Anschlagdruck bei der Betätigung der einzelnen Tasten mit berücksichtigt werden. Aus den  
 5 einzelnen Kurven (für jede Taste eine), die die auf die jeweilige Taste ausgeübte Kraft in ihrem zeitlichen Verlauf beschreiben, können einzelne skalare Maßzahlen für den Druck berechnet werden. Denkbar wären hier beispielsweise  
 10 Mittelwert, Maximalwert oder Integral über die Kurve jeweils bezogen auf einen einzelnen Anschlag. Am einfachsten ist es, sich hier auf eine derartige Maßzahl zu beschränken, deren Wert jeweils beim Loslassen einer Taste zu berechnen und ihn schließlich als vierte Komponente in  
 15 den Beobachtungsvektor mit aufzunehmen. In diesem Fall könnte das Beobachtungsfeld unseres Beispiels folgende Gestalt annehmen:

Ereignisart	↓	↓	↑	↑	↓	↑	↓	↑	↓	↑
Taste	linkes_Shift	h	h	linkes_Shift	u	u	n	n	d	d
Zeitpunkt	0,000	0,024	0,257	0,301	0,388	0,522	0,621	0,698	0,984	1,131
Druckmaßzahl	-	-	23,5	55,6	-	19,8	-	50,3	-	60,3

25 Einschränkungen

Es ist auch denkbar, nur eine Teilmenge der angeführten Meßgrößen zu erfassen.  
 Läßt man z.B. die Ereignisart weg (das impliziert, daß dann  
 30 nur die Anschlagzeitpunkte, nicht aber die Loslaßzeitpunkte gemessen werden), so stützen sich alle weiteren Analysen auf die Übergangsdauern von einem Tastenanschlag zum nächsten.

Mißt man hingegen nur die Ereignisart und die Taste, so kann man lediglich die Tastenwahl und die Reihenfolge der beobachteten Elementarereignisse betrachten. Untersuchungen haben ergeben, daß darin bereits einige sehr  
5 personentypische Merkmale enthalten sind. Bei genügend langen Texten lassen sich auch ohne die Messung der Zeitpunkte Identitätstests auf einem hohen Sicherheitsniveau durchführen. Dies kommt auch dadurch zum Ausdruck, daß viele der weiter unten beschriebenen  
10 Kenngrößen die Zeit überhaupt nicht mit berücksichtigen. Zieht man den Anschlagdruck ebenfalls mit hinzu, so ergeben sich noch einige weitere sinnvolle Meßgrößen-Kombinationen.

#### Technische Geräte

15 Die hier vorgestellte Idee kann mit Hilfe der unterschiedlichsten technischen Geräte realisiert werden. Verzichtet man auf die Druckmessung, so könnte man für etliche Anwendungen beispielsweise ganz auf Standardgeräte  
20 der Computertechnologie zurückgreifen. Denkbar sind aber auch speziell angefertigte Geräte, bei denen z.B. die Zeit- und/oder Druckmessung in der Tastatur integriert sind. Diese können evtl. zusätzlich noch mit Speicher- und/oder Rechenfähigkeit ausgestattet sein. Des weiteren erscheint  
25 für einige Anwendungen die Einbeziehung von Mikroprozessorchipkarten sinnvoll zu sein. Auf ihnen könnte man beispielsweise die personentypischen Muster speichern und evtl. sogar die Identifizierung selbst durchführen.

#### 30 Analyse der Meßdaten

Als Rohdaten stehen dem System die oben beschriebenen Beobachtungsfelder zur Verfügung. Wie diese weiter

analysiert werden können, um schließlich personentypische Merkmale zu erhalten, ist Thema dieses Kapitels.

### Kenngrößen

5

Um das Tippverhalten einer Person zu analysieren, kann man aus den aufgenommenen Beobachtungen eine Reihe verschiedener Kenngrößen berechnen. Prinzipiell ist eine Kenngröße einfach ein, mit Hilfe einer vorgegebenen

10 Rechenvorschrift aus einem Beobachtungsfeld berechneter Wert. Es sind somit beliebig viele verschiedene Kenngrößen denkbar. Sinnvollerweise versucht man sich jedoch auf solche zu beschränken, die möglichst viel über das Tippverhalten einer Person aussagen.

15

Ein Beispiel für eine Kenngröße wäre etwa die Anzahl, wie oft der Schreiber in dem Beobachtungszeitraum die Taste "s" gedrückt hat. Diese Größe ist zwar sicherlich sehr einfach aus dem Beobachtungsfeld zu ermitteln, dürfte jedoch im

20 allgemeinen nur äußerst wenig über die Identität des Schreibers aussagen. Im folgenden werden etliche Beispiele für Kenngrößen angeführt, die jeweils einen gewissen Aspekt des Tippverhaltens untersuchen und auch tatsächlich personentypisch sein können.

25

### Gebrauch der Shifttasten, Tastenwahl

Aus dem Beobachtungsfeld kann leicht ermittelt werden, wie oft der Schreiber die linke und wie oft die rechte

30 Shifttaste benutzt hat, um beispielsweise ein großes 'A' einzugeben. Eine mögliche Kennzahl wäre somit die relative Häufigkeit des Gebrauchs der linken Shifttaste bei der Eingabe eines großen 'A'.

Konkretes Beispiel:

In dem Beobachtungsfeld tritt einmal der Fall "'A' mit der linken Shifttaste geschrieben" und neunmal der Fall "'A' mir der rechten Shifttaste geschrieben" auf. Der Wert der Kenngröße ist dann  $K = 10\%$ .

Entsprechend lassen sich für alle anderen Tasten, mit denen durch Großschreibung eine Eingabe getätigt werden kann, analoge Kenngrößen ermitteln. Zusammengefasst charakterisieren sie das Verhalten des Schreibers bei der Großschreibung.

Auf dieselbe Art lassen sich auch Kennzahlen für andere Tasten mit Wahlmöglichkeit berechnen. So kann beispielsweise auf einer gewöhnlichen Computertastatur eine '1' sowohl auf dem Haupttastaturfeld als auch auf dem Nummernfeld eingegeben werden. Eine mögliche Kennzahl für die Auswahl der '1' ist wieder die beobachtete relative Häufigkeit. Hierbei sollten jedoch einzeln auftretende Ziffern und längere Zahlenkolonnen unterschieden werden.

Fehleranfälligkeit

Ebenfalls peronentypisch ist die Anzahl der Tippfehler bezogen auf die Länge des eingegebenen Textes. Eine einfach zu bestimmende Kenngröße hierfür ist beispielsweise die relative Häufigkeit der Zeichen 'Backspace' und 'Delete' in Bezug auf die Gesamtzahl der eingegebenen Zeichen. Ist dem System der dem Beobachtungsfeld zugrunde liegende Vorlagentext bekannt, so kann es auch die tatsächliche Häufigkeit der gemachten Schreibfehler berechnen. Zusammen mit obiger auf dem Auftreten der Korrekturtasten beruhender Kenngröße ließe sich schließlich sogar die Zahl der vom

Schreiber nicht entdeckten Tippfehler ermitteln. All diese Kenngrößen enthalten eine bestimmte Information über den Aspekt "Fehleranfälligkeit" des Schreibers.

## 5 Verschränkungen

In den allermeisten Fällen sind die Benutzer des Systems zumindest einigermaßen an den Umgang mit der Tastatur gewohnt. Sie müssen demnach nicht jede Taste von neuem  
10 suchen, um diese dann zu drücken und anschließend die nächste Taste wieder zu suchen. Vielmehr entsteht ein gewisser Schreibfluß. Dabei kann es sehr leicht vorkommen, daß man die zweite Taste drückt, noch bevor man die erste losgelassen hat. Beispielsweise ist mit jeder der beiden  
15 unten angegebenen Folgen von Elementarereignissen die korrekte Eingabe des Wortes "er" verbunden:

Folge 1 (ohne Verschränkung): e↓ e↑ r↓ r↑

Folge 2 (mit Verschränkung): e↓ r↓ e↑ r↑

20 Bei der zweiten Folge tritt allerdings eine sogenannte Verschränkung auf, d.h. das Drücken der Taste 'r' erfolgt in diesem Fall noch bevor die Taste 'e' wieder losgelassen wurde. Trotzdem wird offensichtlich auch hier das Wort "er" eingegeben, da der Druck der Taste 'e' vor dem der Taste  
25 'r' erfolgt.

Untersuchungen haben gezeigt, daß das Auftreten von Verschränkungen in hohem Maße personentypisch ist. Sie kommen in der Regel bei jedem vor, der einigermaßen an den  
30 Umgang mit einer Tastatur gewohnt ist; die jeweiligen Häufigkeiten für die verschiedenen Tastenkombinationen sind jedoch von Person zu Person unterschiedlich.

Eine für die Untersuchung von Verschränkungen geeignete Kenngröße ist die relative Häufigkeit mit der beispielsweise die Tastenkombination 'e''r' ohne Verschränkung in dem Beobachtungsfeld auftritt. Die  
5 Bezugsgröße ist dabei die Gesamtzahl der beobachteten 'e''r'-Kombinationen. In entsprechender Weise kann auch für alle anderen Tastenkombinationen jeweils die analoge Kenngröße ermittelt werden. Bereits in den Kennzahlen zu den dreißig bis fünfzig häufigsten Kombinationen steckt ein  
10 sehr hohes Maß an personenidentifizierender Information.

#### Anschlagdauer

Ebenfalls charakteristisch für eine Person ist die Zeit,  
15 wie lange sie die einzelnen Tasten gedrückt hält. Diese hängt nicht unbedingt von der gesamten Schreibgeschwindigkeit der Person ab, sondern ist deutlich von dieser zu trennen. Eine eher langsam schreibende Person kann auf einzelnen Tasten eine deutlich geringere  
20 durchschnittliche Anschlagdauer haben, als ein äußerst schneller Schreiber.

Im folgenden soll die Ermittlung von geeigneten Kennzahlen am Beispiel der Anschlagdauer der Taste 'e' näher erläutert werden. Wurde die Taste 'e' im Verlauf der Messung  
25 beispielsweise 100 mal gedrückt, so können aus dem Beobachtungsfeld leicht die 100 zugehörigen Anschlagdauern sowie die zugehörige empirische Verteilungsfunktion ermittelt werden. Diese Verteilungsfunktion kann mit Hilfe eines geeigneten statistischen Modells durch eine  
30 parametrisierte Verteilungsfunktion angenähert werden. Die Schätzer der zugehörigen Verteilungsparameter lassen sich analytisch aus den gegebenen 100 beobachteten Anschlagdauern berechnen. Sie bilden die die Anschlagdauer von 'e' beschreibenden Kenngrößen.

Konkretes Beispiel:

5 Nimmt man als statistisches Modell an, daß die  
Anschlagdauer der Taste 'e' mit den Parametern  $\mu$  und  $\sigma^2$   
normalverteilt ist, so erhält man für 'e' zwei Kenngrößen:  
Die statistischen Standardschätzer für  $\mu$  und  $\sigma^2$ . Diese sind  
der Mittelwert und die empirische Varianz.  
In entsprechender Weise lassen sich die analogen Kennzahlen  
10 auch für alle anderen Tasten ermitteln. Dabei ist es  
sinnvoll, immer dasselbe statistische Modell zugrunde zu  
legen. Dieses wird in der Regel etwas komplizierter als das  
im Beispiel angegebene sein, was jedoch an der  
prinzipiellen Vorgehensweise nichts ändert.  
15 Die Gesamtheit dieser Kennzahlen für alle Tasten  
beschreiben den Aspekt "Anschlagdauer" des Tippverhaltens  
des Schreibers. Sie enthalten ebenfalls ein hohes Maß an  
personentypischer Information.

20 Übergangsdauer

In ähnlicher Weise wie die Anschlagdauern der einzelnen  
Tasten lassen sich auch die Übergangsdauern der  
verschiedenen Tastenkombinationen analysieren. Mit  
25 Übergangsdauer ist dabei die Zeitspanne vom Drücken einer  
Taste bis zum Drücken der darauffolgenden Taste gemeint.  
Sie hängt im Gegensatz zur Anschlagdauer direkt mit der  
Schreibgeschwindigkeit einer Person zusammen. Ein schneller  
Schreiber hat in der Regel zumindest bei den häufigeren  
30 Tastenkombinationen wesentlich kürzere mittlere  
Übergangsdauern als ein langsamer Schreiber. Die  
Unterschiede von Person zu Person sind jedoch noch  
wesentlich vielschichtiger: Auch zwei in der gesamten  
Schreibgeschwindigkeit sehr ähnliche Benutzer können, wenn

man die einzelnen Tastenkombinationen betrachtet, sehr unterschiedliche Übergangsdauern aufweisen. Je nach Veranlagung fallen dem einen gewisse Anschlagfolgen wesentlich leichter oder sind besser eingeschliffen als bei  
5 dem anderen und umgekehrt.

Wie man nun zu geeigneten, die Übergangsdauern charakterisierenden Kenngrößen kommt, soll im folgenden beispielhaft an der Tastenkombination 'e''r' erläutert  
10 werden. Die prinzipielle Vorgehensweise ist wieder die gleiche wie bei den Anschlagdauern. Wurden beispielsweise im Verlauf der Messung die Tasten 'e' und 'r' 20 mal direkt hintereinander in dieser Reihenfolge gedrückt, so bilden die zugehörigen 20 Übergangsdauern die  
15 Berechnungsgrundlage. Sie können aus dem Beobachtungsfeld leicht ermittelt werden. Man versucht nun wieder mit Hilfe eines geeigneten statistischen Modells die empirische Verteilungsfunktion der beobachteten Zeiten durch eine parametrische anzunähern. Die Schätzwerte der zugehörigen  
20 Verteilungsparameter bilden dann die Kenngrößen. Das hier zugrunde gelegte statistische Modell kann durchaus von dem für die Analyse der Anschlagdauern verwendeten abweichen. Wieder lassen sich in entsprechender Weise auch die  
25 analogen Kennzahlen für alle anderen Tastenkombinationen ermitteln. Sinnvollerweise wird man sich jedoch auf die wichtigsten Kombinationen beschränken. Die Gesamtheit der entsprechenden Kennzahlen beschreiben schließlich den Aspekt "Übergangsdauer" des Tippverhaltens des Schreibers.  
Anschlagdruck

30

Wird bei den durchgeführten Messungen der Tastendruck ebenfalls mit erfaßt, so lassen sich auch hieraus eine Reihen von Kenngrößen berechnen. Die Vorgehensweise ist dabei zu der bei den Anschlagdauern identisch. Es werden



wieder die einzelnen Tasten getrennt voneinander betrachtet und die zugehörigen empirischen Verteilungsfunktionen der beobachteten Druckmessungen durch parametrische Verteilungen angenähert. Der einzige Unterschied besteht im Prinzip in dem zugrunde gelegten statistischen Modell. Außerdem sind, falls verschiedene Druckmeßgrößen erfaßt worden sind (z.B. Mittelwert und Maximalwert), diese auch getrennt voneinander zu analysieren.

- 5  
10 Wieder enthält die Gesamtheit der so ermittelten Kenngrößen charakteristische Informationen über das "Anschlagdruck-Verhalten" des Schreibers.

#### Anlaufphase und Prägungen

- 15 Wie anhand dieser sechs Beispiel näher erläutert wurde, lassen sich aus einem Beobachtungsfeld eine ganze Reihe unterschiedlicher Kenngrößen berechnen. Jede einzelne von ihnen sagt etwas über die verschiedenen Eigenschaften des Schreibers aus, was dessen Tippverhalten angeht. Neben den so, aus dem gesamten Beobachtungsfeld gewonnenen Charakteristika gibt es noch andere Aspekte des Tippverhaltens, die in gewisser Weise eine Ebene höher anzusiedeln sind. Zwei Beispiele hierfür sind die  
20 Anlaufphase und die sogenannten Prägungen; sie werden im  
25 folgenden kurz erläutert.

- Beginnt man nach einer kurzen Pause (es genügen wenige Minuten, in denen man nichts getippt hat) damit, einen  
30 längeren Text einzugeben, so dauert es für gewöhnlich eine gewisse Zeit, bis man in seinen normalen Schreibfluß gefunden hat. Diese Zeitspanne wird hier als Anlaufphase bezeichnet. Sie ist in der Regel durch eine erhöhte Fehleranfälligkeit und einen ungleichmäßigeren

Schreibrhythmus gekennzeichnet. Die gesamte Schreibgeschwindigkeit ist nicht notwendigerweise niedriger. Sowohl die Dauer der Anlaufphase, als auch deren Ausprägungsgrad und die Art der einzelnen Ausprägungen sind  
5 weitere personentypische Aspekte des Tippverhaltens.

Als Prägungen werden im folgenden einzelne Wörter und Tastenfolgen bezeichnet, die von einer Person besonders häufig eingegeben werden. Beispiele hierfür sind der Name  
10 und die Kennung der Person, sowie evtl. spezielle Kommandos oder Syntaxelemente einer Programmiersprache. Die Prägungen sind somit im allgemeinen von Person zu Person unterschiedlich.

15 Sie sind für die Analyse des Tippverhaltens von besonderem Interesse, weil sie in der Regel anders geschrieben werden, als die entsprechenden einzelnen Zeichen bzw. Zeichenkombinationen. Sie müssen demnach eigenständig analysiert werden, um zum einen nicht die Kenngrößen für  
20 den "normalen Text" zu verfälschen und um zum anderen die in ihnen steckende zusätzliche Information ebenfalls zu berücksichtigen.

Warum diese beiden Aspekte, die Anlaufphase und die  
25 Prägungen, eine Ebene über den oben angeführten Eigenschaften anzusiedeln sind, wird ersichtlich, wenn man die zugehörigen Kenngrößen betrachtet. Sowohl die Anlaufphase als auch die Prägungen lassen sich nämlich ebenfalls durch die oben beschriebenen Kenngrößen  
30 charakterisieren. Der einzige Unterschied besteht darin, daß man nun nicht das gesamte Beobachtungsfeld, sondern lediglich Teile davon zu deren Berechnung heranzieht. So werden beispielsweise als Grundlage für die Ermittlung der die Prägung "printf" beschreibenden Kenngrößen lediglich

diejenigen Beobachtungen verwendet, die auch in direktem Zusammenhang mit dem Schreiben dieser Zeichenfolge stehen. Entsprechend könnte man für die Anlaufphase z.B. nur die ersten 300 Anschläge berücksichtigen.

5

Insgesamt erhält man also auch für derartige höherstehende Aspekte des Tippverhaltens eine Reihe von charakterisierenden Kenngrößen. Dabei kann auf die grundlegenden Ideen zurückgegriffen werden.

10

#### Aspekte des Tippverhaltens

In den vorangegangenen Abschnitten wurde anhand etlicher Beispiele erläutert, wie man die einzelnen Aspekte des Tippverhaltens mit Hilfe verschiedener Kenngrößen analysieren kann. Dabei sind nahezu beliebig viele unterschiedliche Berechnungsvorschriften zur Gewinnung sinnvoller Kenngrößen denkbar, die jedoch alle eine Gemeinsamkeit aufweisen: Die erhaltenen Größen charakterisieren jede für sich irgendeine Eigenart im Tippverhalten des Schreibers. An dieser Stelle werden deshalb nochmals die wichtigsten Aspekte des Tippverhaltens aufgeführt.

25

1. Schreibrhythmus (zeitlicher Aspekt)

\* Anschlagdauern

\* Übergangsdauern

2. Anschlagdruck

3. Anschlagfolge

30

\* Tastenauswahl (z.B. Shifttastengebrauch)

\* Verschränkungen

4. Fehlerverhalten

\* Häufigkeiten (bemerkte und unbemerkte Tippfehler)

- \* personenspezifische Tippfehler (z.B. Buchstabendreher, spezielle Wörter, ...)

- \* Korrekturverhalten

#### 5. Anlaufphase

- 5 \* Ausprägungsgrad
- \* Art der Ausprägungen
- \* Dauer der Anlaufphase

#### 6. Prägungen

- 10 Insgesamt muß man sich schließlich für eine gewisse Anzahl an Kenngrößen entscheiden (mindestens eine, in der Regel jedoch etliche Hundert). Zusammengenommen werden diese im folgenden immer als Kenngrößenvektor bezeichnet.

#### 15 Muster

- In einem letzten Schritt kann man nun aus dem Kenngrößenvektor das personentypische Muster ermitteln, das im weiteren auch als Referenzmuster für diese Person oder als psychometrisches Muster dieser Person bezeichnet wird. Dieses Muster besteht aus dem besagten Kenngrößenvektor und einem zusätzlichen Gewichtsvektor derselben Dimension. Es wird also lediglich jede Kenngröße noch mit einem bestimmten Gewicht, das zwischen Null und einem festen
- 20
- 25 Höchstwert liegt, versehen.

- Die der Musterberechnung zugrunde liegende Idee ist die folgende: Betrachtet man das Tippverhalten einer Person, so wird es in der Regel zum großen Teil mit dem vieler anderer übereinstimmen. So benutzt beispielsweise jeder gelernte Schreibmaschinenschreiber grundsätzlich die rechte
- 30
- Shifttaste, um ein großes 'A' einzugeben. Andererseits hat aber auch jede Person gewisse Eigenarten, durch die sie sich vom Schreibverhalten anderer unterscheidet. Ziel der

Musterberechnung ist es nun, gerade diese charakteristischen Eigenschaften ausfindig zu machen und die zugehörigen Kenngrößen schließlich mit entsprechend hohen Gewichten zu versehen. Dadurch wird die spätere Identifizierung eines Berechtigten bzw. entsprechend die Ablehnung eines Unberechtigten erheblich erleichtert. Die drei wichtigsten Methoden, mit deren Hilfe die charakteristischen Merkmale erkannt werden können, sind die folgenden.

10

1. Vergleich mit dem empirisch ermittelten "Durchschnittsschreibverhalten"

15

Durch Analyse des Schreibverhaltens einer größeren Zahl von Personen lassen sich leicht Standardwerte für die einzelnen Kenngrößen ermitteln, d.h. Werte oder sogar Intervalle für die betrachteten Größen, die das Tippverhalten der meisten Probanden widerspiegeln. Liegt nun eine Kenngröße des gerade untersuchten Vektors außerhalb des entsprechenden Intervalls bzw. relativ weit von dem entsprechenden Normalwert entfernt, so deutet dies auf eine charakteristische Eigenart hin. Einer derartigen Kenngröße sollte dementsprechend ein großes Gewicht zugeordnet werden.

20

2. Vergleich mit den Kenngrößenvektoren der anderen dem System bekannten Personen

25

Liegen dem System bereits einige Kenngrößenvektoren anderer Personen vor, so lassen sich mit deren Hilfe ebenfalls einige Eigenarten im Schreibverhalten der augenblicklich untersuchten Person feststellen. Weicht eine Kenngröße von allen oder zumindest den meisten entsprechenden Werten der anderen Personen stark ab, so deutet dies wieder auf ein personentypisches Merkmal hin. Auf diese Art und Weise kann eine besonders hohe

30

Trennschärfe zwischen den dem System bekannten Personen erreicht werden.

3. Berechnung der Varianzen der einzelnen Kenngrößen

Die dritte hier vorgestellte Methode bezieht sich nicht  
5 auf Vergleichswerte, sondern betrachtet lediglich die  
Konstanz im Schreibverhalten der untersuchten Person.  
Demnach ist es sinnvoll, eine Kenngröße, die nur sehr  
geringen Schwankungen unterworfen ist, wesentlich höher  
zu gewichten als eine, die bei dem Schreiber mal diesen  
10 und mal jenen Wert annehmen kann. Ein Maß für diese  
Konstanz bieten die im Zusammenhang mit der Berechnung  
der einzelnen Kenngrößen auftretenden Varianzen.

Gesamtablauf

15 In diesem Kapitel wird beschrieben, wie sich die bisher  
vorgestellten Methoden und Konzepte zu einer geschlossenen  
Anwendung zusammenfügen lassen. Als Grundlage dazu dient  
weiterhin das angegebene einführende Beispiel.

20 Lernvorgang

Während der Benutzer den vorgegebenen Text abtippt, führt  
das System die oben beschriebenen Messungen durch und  
25 erhält dadurch das Beobachtungsfeld. Aus diesem werden dann  
die vorher festgelegten Kenngrößen ermittelt. Liegen dabei  
für eine Größe keine oder zu wenig Beobachtungen vor, so  
wird diese übergangen und im weiteren nicht mehr  
berücksichtigt. Aus dem so erhaltenen Kenngrößenvektor wird  
30 schließlich, z.B. mit Hilfe einer der oben beschriebenen  
Methoden oder einer Kombination davon, das Referenzmuster  
berechnet.

Was nun im einzelnen alles vom System dauerhaft abgespeichert werden muß, hängt u.a. von dem verwendeten Identifizierungsverfahren ab. Es kann das gesamte Beobachtungsfeld, der Kenngrößenvektor oder das

5 Referenzmuster sowie jegliche Kombination dieser drei Datensätze gespeichert werden. Im einfachsten Fall genügt es jedoch, sich das Referenzmuster zu merken.

Neben einem solch einfachen Ablauf für das Lernverfahren sind noch viele andere Möglichkeiten denkbar. Insbesondere

10 kann es sinnvoll sein, den gesamten Lernvorgang adaptiv zu gestalten. Dazu könnte man explizit immer wieder einen neuen Lerndurchgang anstoßen oder implizit die bei der Identifizierung gewonnenen Beobachtungsfelder verwenden. Dabei dürften natürlich nur diejenigen Versuche

15 Berücksichtigung finden, bei denen der Benutzer eindeutig vom System identifiziert werden konnte. Einen völlig anderen Aspekt der Adaptivität erhält man, wenn man immer nach der Aufnahme eines neuen Benutzers die Gewichtungen der Kenngrößenvektoren aller anderen Benutzer neu

20 durchführt. Dadurch ließe sich die Trennschärfe innerhalb der Benutzergruppe optimieren.

#### Identifizierung

25 Das eigentliche Ziel der gesamten Anwendung ist die Identifizierung. Das System soll auf einem bestimmten Sicherheitsniveau dazu in der Lage sein, berechnete Anmelder von unberechneten anhand ihres Tippverhaltens zu unterscheiden.

30

Der erste Schritt des Identifikationsvorgangs ist das Erfassen der verschiedenen Meßgrößen, während der Anmelder A den Testtext abtippt. Dadurch erhält das System ein Beobachtungsfeld  $F_A$ . Für die anliegende Entscheidung, ob A

mit dem Behaupteten B identisch ist, stehen dem System die von ihm im Rahmen des Lernvorgangs von B abgespeicherten Daten zur Verfügung: das Beobachtungsfeld  $F_B$ , der Kenngrößenvektor  $K_B$  sowie das Referenzmuster  $M_B$ .  
5 Für die Vorgehensweise bei der letztendlichen Entscheidungsfindung sind nun einige, grundsätzlich verschiedene Ansätze denkbar, die evtl. auch miteinander kombiniert werden können. Im folgenden werden die drei wichtigsten Methoden näher erläutert:

10

#### 1. Direkter Mustervergleich

Das naheliegendste und einfachste Verfahren besteht darin, aus dem Beobachtungsfeld  $F_A$  in zum Lernvorgang analoger Weise einen Kenngrößenvektor  $K_A$  zu berechnen und diesen mit dem Muster  $M_B$  zu vergleichen. Der Vergleich könnte beispielsweise durch die Berechnung des gewichteten euklidischen Abstands der beiden Kenngrößenvektoren  $K_A$  und  $K_A$  erfolgen (gewichtet bedeutet dabei eine durch die Komponenten des Gewichtsvektors festgelegte lineare Skalierung der einzelnen Dimensionen). Der Nachteil dieses Verfahrens besteht in der dazu nötigen Mindestlänge des Testtextes. Ist dieser nämlich zu kurz, so können aus dem zugehörigen Beobachtungsfeld viele der Kenngrößen nicht mehr oder nur unzureichend berechnet werden. Ein  
15  
20  
25 sinnvoller Vergleich ist dann nicht mehr möglich.

#### 2. Vergleich des Beobachtungsfeldes $F_A$ mit dem Referenzmuster $M_B$

30 Dieses Verfahren kann auch bei sehr kurzen Testtextlängen noch gut eingesetzt werden, da es auf die Berechnung der verschiedenen Kenngrößen aus dem Beobachtungsfeld  $F_A$  verzichtet. Statt dessen wird das Beobachtungsfeld  $F_A$  direkt mit dem Muster  $M_B$



5 verglichen. Es wird eine Art Wahrscheinlichkeitsmaß  
dafür berechnet, daß diejenige Person, deren  
Schreibverhalten durch  $M_B$  charakterisiert ist, die  
gemessenen Beobachtungen  $F_A$  hervorruft. Für einen  
Berechtigten wird diese Maßzahl in der Regel relativ  
hoch, für einen Unberechtigten relativ niedrig  
ausfallen. Durch das geforderte Sicherheitsniveau wird  
schließlich der Schwellenwert für die letztendliche  
Entscheidung festgelegt.

10

### 3. Direkter Vergleich der Beobachtungsfelder

Die letzte hier vorgestellte Methode verzichtet auf  
jeglichen theoretischen Hintergrund. Mit Hilfe nicht-  
klassischer Ansätze, wie z.B. einem neuronalen Netz,  
15 ist es durchaus möglich, die beiden Beobachtungsfelder  
 $F_A$  und  $F_B$  in gewisser Weise direkt miteinander zu  
vergleichen. Eines der Hauptprobleme, die variable  
Anzahl der Einzelbeobachtungen in  $F_A$  und  $F_B$  auf eine  
feste Zahl an Eingabeneuronen abzubilden, läßt sich  
20 beispielsweise durch die sogenannte topologische  
Codierung lösen. Die mit Hilfe eines derartigen  
Verfahrens erzielten Ergebnisse hängen sehr stark von  
den verwendeten Lernalgorithmen, sowie von den zugrunde  
liegenden Lerndaten ab. Dennoch lassen sich sicherlich  
25 auch auf dieser Idee, die Beobachtungsfelder direkt  
miteinander zu vergleichen, basierende  
Identifikationssysteme realisieren.

30 Diese drei kurzen Absätze sollen lediglich verdeutlichen,  
daß durchaus mehrere Verfahren, die sich grundlegend  
voneinander unterscheiden, für die Realisierung des  
Identifikationsverfahrens denkbar sind. Entscheidend für  
die letztendlich erzielte Trennschärfe der Implementierung

sowie für die geforderten Mindestlängen der Lern- bzw. Testtexte sind sicherlich die folgenden Punkte:

- \* Verwendete Meßgrößen
- 5 \* Analyisierte Aspekte des Tippverhaltens
- \* Verwendete Kenngrößen
- \* Verwendetes Identifikationsverfahren

10 Exaktere Beschreibung der Unterschiede zum Stand der Technik

Analyse des Tippverhaltens

- 15 Wie in dem Kapitel "Stand der Technik" bereits erwähnt, stammt die dem hier vorgestellten System ähnlichste Methode von S. J. Shepherd. Die hier in dem Kapitel Lösung niedergelegten Ansätze sind jedoch sowohl von den zugrunde liegenden Ideen, als auch von den verwendeten Verfahren her
- 20 wesentlich umfassender. Die wichtigsten Unterschiede werden im folgenden kurz erläutert.

- \* Analyisierte Aspekte des Tippverhaltens
- Das System von Shepherd zielt ausschließlich auf die
- 25 Schreibdynamik und den Tastendruck ab. Er erwähnt als mögliche Merkmale lediglich die zeitlichen Aspekte Übergangsdauer ("Intervals between keystrokes"), Anschlagdauer ("Duration of keystrokes"), Tipptrate ("Rate of typing"; damit bezeichnet er die Anzahl der
- 30 geschriebenen Zeichen oder Wörter pro Minute) und Fehlerhäufigkeit ("Frequency of errors"), sowie den Tastendruck ("Force of keystrokes"). Alle anderen Aspekte - insbesondere die Tastenauswahl und Verschränkungen (siehe Abschnitt Aspekte des

Tippverhaltens) - wurden außer Acht gelassen.  
Er geht zwar kurz darauf ein, daß Verschränkungen  
("rollover") vorkommen und von der Tastatur in  
gewohnter Weise behandelt werden, verwendet die  
5 diesbezüglichen Beobachtungen jedoch lediglich dazu,  
entsprechende Fehler bei der Berechnung der Anschlag-  
und Übergangsdauern zu vermeiden. D.h. er beschränkt  
sich hier einzig und allein auf den zeitlichen Aspekt  
und wertet die Reihenfolgeinformation an sich nicht  
10 weiter aus.  
Darüber hinaus erwähnt Shepherd noch die Möglichkeit,  
durch die von der Tastatur gelieferten Scan codes auch  
Tasten voneinander unterscheiden zu können, die  
mehrfach vorhanden sind (wie z.B. die beiden Shift-,  
15 CTRL- und ALT-Tasten, sofern vorhanden). Hierbei stellt  
er jedoch ebenfalls lediglich die Bedeutung für die  
zeitlichen Aspekte heraus. Auf die Tastenauswahl an  
sich, wie sie hier im Abschnitt "Analyse der Meßdaten"  
beschrieben ist, geht er nicht ein.  
20

\* Verwendete Kenngrößen

Die von Shepherd angegebenen Kenngrößen beschränken  
sich auf die Charakterisierung der Aspekte Anschlag-  
25 und Übergangsdauer, wobei zwischen den verschiedenen  
Tasten keine Unterschiede gemacht werden. Er verwendet  
lediglich Mittelwerte und Varianzen.

\* Verwendete Verfahren

Das von Shepherd dargelegte Identifikationsverfahren  
30 entspricht dem hier erläuterten "direktem  
Mustervergleich" (siehe Punkt 1. im Abschnitt  
Identifizierung). Insbesondere wird die wesentlich  
kompliziertere und diesem Verfahren weit überlegene  
Methode des Vergleichs des Beobachtungsfeldes mit dem

Referenzmuster (siehe Punkt 2. Im Abschnitt Identifizierung) nicht erwähnt. Auch die den hier vorgestellten Kenngrößen zugrunde liegenden Verfahren fehlen im alten Vorschlag vollkommen.

5

Abschließend sei noch erwähnt, daß es durchaus möglich ist, ein Identifikationsverfahren zu realisieren, das alleinig auf neuen hier dargelegten Ideen basiert. Zudem lassen sich offensichtlich die neuen und die alten Ansätze hervorragend miteinander kombinieren.

10

#### Analyse der Mausbewegungen

Im Gegensatz zur Unterschriftenerkennung wird bei der im Abschnitt "Anwendungsgebiet" angeführten Analyse des Umgangs mit der Maus nicht die Bewegung beim "Schreiben" eines bestimmten Wortes, sondern die Geschicklichkeit des Benutzers ganz allgemein betrachtet. Es werden die bei der Bedienung des Systems anfallenden Mausbewegungen zusammen mit den getätigten Mausklicks untersucht. Diese Daten lassen sich mit ähnlichen Methoden wie beim Tippverhalten mit einem zum jeweiligen Benutzer hinterlegten Referenzmuster vergleichen. Charakterisierende Merkmale hierbei sind z.B. das Verhalten beim Ansteuern eines anzuklickenden Symbols oder der Umgang mit Pull-down-Menüs oder Scroll-bars. Es werden somit ganz allgemeine Kennzeichen untersucht und nicht das Verhalten in einer ganz speziellen, genau vorgegebenen Situation, wie dies bei der Unterschriftenerkennung der Fall ist.

30

## A n s p r ü c h e :

5

1. Verfahren zur Verifizierung der Identität eines  
Benutzers einer mit einer Tastatur zur Erzeugung  
10 alphanumerischer Zeichen zu bedienenden  
Datenverarbeitungsanlage, wobei aus dem Muster von  
aktuellen Beobachtungen an der Tastatur mindestens  
ein auf das Tippverhalten des Benutzers hinweisender  
Kenngrößenvektor abgeleitet und dieser  
15 Kenngrößenvektor mit dem Kenngrößenvektor verglichen  
wird, der aus einem in der Datenverarbeitungsanlage  
abgelegten Referenzmuster abgeleitet ist,  
d a d u r c h g e k e n n z e i c h n e t ,  
daß unabhängig von der eingegebenen Zeichenfolge der  
20 Kenngrößenvektor aus der Reihenfolge der Aktionen  
Drücken bzw. Loslassen aufeinanderfolgend betätigter  
Tasten in Verbindung mit den Beobachtungen betreffend  
die Auswahl der betätigten Tasten ermittelt wird.
- 25 2. Verfahren nach Anspruch 1 ,  
d a d u r c h g e k e n n z e i c h n e t ,  
daß zusätzlich zu den Beobachtungen als Meßgrößen die  
Zeitpunkte der Ereignisse bzw. Betätigungen erfaßt  
werden.
- 30 3. Verfahren nach einem der vorhergehenden Ansprüche,  
d a d u r c h g e k e n n z e i c h n e t ,  
daß zusätzlich zu den Beobachtungen als Meßgröße der  
Tastendruck erfaßt wird.

4. Verfahren nach einem der vorhergehenden Ansprüche,  
d a d u r c h g e k e n n z e i c h n e t ,  
daß aus dem Feld der Beobachtungen und/oder Meßgrößen  
5 Kenngrößen errechnet werden, die das Tippverhalten  
wiederspiegeln.
5. Verfahren nach einem der vorhergehenden Ansprüche,  
d a d u r c h g e k e n n z e i c h n e t ,  
10 daß zusätzlich zu den Beobachtungen als Meßgröße die  
Anschlagsdauer einer Taste erfaßt wird.
6. Verfahren nach Anspruch 5,  
d a d u r c h g e k e n n z e i c h n e t ,  
15 daß mehrere verschiedene Kenngrößen gewichtet zu  
einem Kenngrößenvektor zusammengefaßt werden.
7. Verfahren nach einem der vorhergehenden Ansprüche,  
d a d u r c h g e k e n n z e i c h n e t ,  
20 daß das Referenzmuster durch Abschreiben einer  
vorgegebenen alphanumerischen Folge erstellt wird.
8. Verfahren nach einem der Ansprüche 1 bis 6,  
d a d u r c h g e k e n n z e i c h n e t ,  
25 daß das Referenzmuster durch freies Schreiben einer  
nicht vorgegebenen alphanumerischen Folge erstellt  
wird.
9. Verfahren nach einem der vorhergehenden Ansprüche,  
d a d u r c h g e k e n n z e i c h n e t ,  
30 daß das Referenzmuster während der Eingabe laufend  
aktualisiert wird.
10. Verfahren zur Verifizierung der Identität eines

Benutzers einer mit einer Maus zu bedienenden  
Datenverarbeitungsanlage,  
dadurch gekennzeichnet,  
daß der aktuelle Bewegungsablauf der Maus auf dem  
5 Maus-Pad in X-Y-Koordination und die  
Bewegungsgeschwindigkeit und/oder die Betätigung der  
Maustasten erfaßt und mit entsprechend  
eingespeicherten Dynamikdaten verglichen wird.